再谈同行评议新范式

李政[1]
（1. 中国科协创新战略研究院，北京 100038）

摘要
[目的]本文提出一个同行评议新范式，推动科技评价制度深化改革先立后破。
[方法]运用介尺度分析方法贯通研究同行评议各环节的知识运动变化过程，进而将问题聚焦在知识确权定价内在机理。
[结果]以显然性评价为例，构建同行评议成本理论模型，提出评价者的内心认可与外显评语间可能存在幂律分布。
[局限]为验证模型推论，收集《心理学报》33 篇评审意见，对其否定性关联词做编码和词频分析，发现真假阳性区别和多轮评审的变化趋势，符合模型预期，但样本数较少。
[结论]同行评议新范式包括成本理论模型和公理化体系，后者含 9 条定义和 3 条公理。
关键词 创新人才；开放科学；国家创新体系；信心；显然
分类号 G3

A Second Thought on A New Paradigm of Peer Review

Zheng Li[1]
(1. National Academy of Innovation Strategy, China Association for Science and Technology, Beijing 100038)

Abstract
[Objective] This paper studied on a new paradigm of peer review to promote further reforms of science and technology evaluation.
[Methods] By using Meso-Scale Analysis, a whole picture of knowledge dynamic was sketched, cutting through each part of a peer review. The internal mechanism of right-value interaction of knowledge was focused on as the key question.
[Results] Taking obviousness evaluation in peer review as an example, a theoretical cost model was then constructed, which presented a deductive idea that degrees of reviewers' inner recognition and performances of their explicit comments might have an objective relation that follows power law.
[Limitations] For empirical study, 33 experts' comments from the journal *Acta Psychologica Sinica* were sampled and coded. The expected difference and trend of negative conjunction distributions were found in word frequency analyses in the comment comparisons between first-round rejection (true positives) and multi-round acceptance (false positives), which in some ways verified the inference. But the sample was quite small.
[Conclusions] The new paradigm contains the cost model, and the axiomatic system including 9 definitions and 3 axioms.
Keyword Innovative Talents, Open Science, National Innovation System, Confidence, Obvious

---

收稿日期：2025-02-05；修回日期：2025-XX-XX
作者简介：李政，男，博士，中国科协创新战略研究院副研究员，研究方向为科技政策、科技与社会等。信箱：li.zheng@cast.org.cn。

再谈同行评议新范式

1. 引言

党和国家高度重视构建支持全面创新体制机制，统筹推进教育科技人才一体改革[1]-[3]。习近平总书记多次强调，“加快建立以创新价值、能力、贡献为导向的科技人才评价体系”“要明确‘破四唯’后怎么‘立’的评价方式和标准”“建立健全责任明晰、链条完整、环环相扣的工作机制”[4]-[6]。中办国办、教育部科技部、地方政府出台多项政策[2][7][8]，多个高校和科研院所依据自身特点积极谋划“破四唯”落地举措，取得了一定成效。然而，“立新标”探索遭遇很多困难[7]，改革已入深水区、无人区。未来，贯彻落实党中央精神、加速提升国家创新体系整体效能、在制度上破立并举、先例后破的关键，就在于评价制度的传承发展，在于对同行评议的改革创新。

关于什么是同行评议，它的对象、内容、形式、结构等等，大家见仁见智，但论及其主要功能，无外乎把控质量、及时纠错、防范虚假、判断价值等。然而，现代同行评议已愈加不能胜任制度设计初衷，被责任分散、人情交易、外行干预等问题蹂躏得千疮百孔[8]-[10]，这些问题的背后正是百年未有之大变局——激烈的国家竞争、迅猛的知识爆炸、频发的交叉创新、冷峻的派系斗争等。同行评议被裹挟其中，被割裂为责任主体不同的多个环节，缺乏有效制度工具实施系统控制，公信力受到威胁，旧范式亟待改变。

本文在学习借鉴现有成果基础上，探讨同行评议新范式。全文包括以下部分：首先回顾专家遴选、指标比对、问答互动等同行评议环节的研究成果，提炼重要变量；随后运用介尺度分析对同行评议各部分做贯通式整合，以便聚焦同行评议中的知识运动变化，提出知识的确权定价机制问题；第三，通过假设演绎法构建成本理论模型，寻找知识权属与权重间的互动关系；接着，收集期刊论文评语做实证分析；最后，总结提炼评价的公理化体系并讨论其推论。不妥之处恳请指正。

2. 文献回顾

2.1. 总系统：基于介尺度的同行评议贯通式系统模型

一般地，同行评议可抽象为由多个要素构成的系统，包含评价主导方[3]（组织方、监督方）、评价方（同行专家、行业机构）、被评方（被评个人、团队、结果、机构、项目、方案、成绩效率等）、标准和流程（基准、指标、权重、参照、程序、规矩等）、评价环境（场地、设备、原则、规章、法律政策、文化风气等）。这些要素贯穿于专家遴选、指标比对、问答互动等多个子系统及介尺度结构中[10]。所谓“介尺度结构”，即介于单元、系统等不同尺度间的、微妙的、过渡性连接结构，分析这类结构的方法就是介尺度分析，它是一套系统思维框架，致力于将复杂系统分成不同层次子系统，考察子系统内单元与其他子系统的关联

---

2 已出台文件包括：《关于分类推进人才评价机制改革的指导意见》《关于深化项目评审、人才评价、机构评估改革的意见》《关于开展清理“唯论文、唯职称、唯学历、唯奖项”专项行动的通知》《关于规范高等学校 SCI 论文相关指标使用 树立正确评价导向的若干意见》《关于深化高等学校教师职称制度改革的指导意见（征求意见稿）》等。

3 有时候，评价主导方与评价方相同。

关系，在稳态之间构建中介、过渡、缓冲区域，揭示其中多要素的竞争和妥协规律[11]-[14]。以同行评议为例，如图 1 所示[4]，位于专家遴选和指标比对之间、指标比对和问答互动之间都存在介尺度；换个更大的视角，指标比对本身也可视为专家遴选与问答互动之间的介尺度。

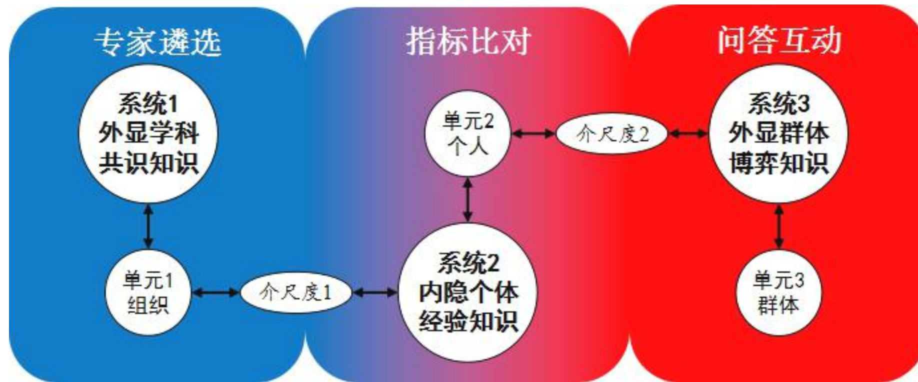


图 1. 介尺度下同行评议系统示意图

2.2. 子系统：专家遴选+指标比对+问答互动

专家遴选涉及非线性整数规划和组合优化，典型理论包括指派模型、二次分配等，通过遗传算法、蚂蚁算法等求得专家遴选的可行解[15][16]。专家遴选的一般过程，是在若干被评对象和同行专家之间建立评审关系，每个对象需要请至少一名专家进行评审，每名专家的评审次数也有上下限，被评对象和专家之间评审关系的紧密度用综合指标来衡量，涉及学科行业匹配度、专家等级等指标。专家遴选的重点，是请最少的同行专家为被评对象建立最紧密的评审关系，重点在于同时协调匹配度（越大越优）、专家总数（越小越优）和每位专家分得对象数（在上下限内），以达到整体效能最优。遗传算法能跳出局部最优但收敛速度较慢，蚂蚁算法收敛速度快却容易陷入局部最优。这些模型虽已应用于基金评审、专题研讨的专家遴选实操中，但在紧密度参数设计及其检验等方面仍可改进。

指标比对涉及多目标多准则决策，常用方法是层次分析法[17]-[21]，通过构建“指标—权重—被评对象”对应关系，完成被评对象相对重要性（层级）排序。该法一般分三步，首先把评价目标分解为多个/层指标，做“指标—指标”两两比较打分，接着做“指标—被评方案”逐一比对打分，最后联立构建“指标—权重—被评对象”矩阵进而计算最终得分；上述三个步骤在打分后均要做一致性检验。该法通过求解高阶正互反矩阵最大特征值的特征向量来检验相对权重的一致性，能控制操作误差、提高评价可靠性；但难点是指标与被评对象的匹配度——在设计指标时存在内生性、非遍历性等问题，在比对打分时不同质的指标均降维成可比的量，在指标失效时又难以检验，故系统误差难以避免。尽管如此，该法因能恰当模拟大多数评价的结构和过程而得到广泛应用，也被我国标准采纳[22]。

问答互动涉及投票模型、非对称信息博弈等，既有模型包括主动观点动力学模型、舆情演化模型等[22]-[27]。这类模型多以贝叶斯定理为基础构建个体及群体观点，通过网络内多主体间信息交互来计算观点变化，能够模拟群体观点与投票结果不相符的悖论情景[28]，揭示出初始群体观点、异议包容度等因素对评价结果

4 这些子系统和介尺度结构并非严丝合缝地衔接在一起，而是分为不同层次和顺序、相互重叠交错、动态变化地存在于同行评议系统中。示意图是对抽象系统的一种形象化展示。

的重要影响，特别是充分的互动讨论对于得到有效评价结果的必要性。理论上，问答互动需要充分的时长、轮次、深度和广度；但实操中，许多评议工作非常短促且隐秘，这有一定合理性——常常是顾及评价工作管理的方便、控制舆情风险、给专家留面子、方便利益交换等——但代价是互动讨论的充分性难以查证或监督，直接导致评议结果丧失公信力和科学性。目前，只有少数同行评议的问答互动能面向社会公开，如《心理学报》期刊、ChinaXiv 预印本平台的部分期刊等[7][29][30]。

综上，同行评议的总模型是多变量运筹模型，主要包含指派分配、层次分析和主动观点三个子系统模型。其中，专家构成、匹配度、评价指标体系、指标权重及其一致性、初始和最终群体观点、投票结果、异议包容度等是重要变量；进一步看，专家构成和指标权重至为关键。

3. 研究结果

3.1. 介尺度下同行评议的焦点

（1） 同行评议的知识

评价的多变量、非线性等特点使得系统模型构建颇为困难，这与评价最基本概念——人的思维——难以实时计量有直接关系。在介尺度分析强调系统主导权竞争和妥协的启发下，本文抓住知识（表征人的思维）作为主要分析视角（表 1），剖析评价过程中知识主导权的变化情况，并把知识笼统地视为事实、观点、推断、猜想、规律等有序信息、逻辑命题的复合体。

①学科共识

评价主导方开启专家遴选时，主要矛盾是评价方与被评方是否匹配，即能否找到掌握评价所需知识的人。此时的知识主要是评价涉及的公理定理、学术常识、基本理论、实验规范、工程惯例、行业历史和前沿热点、国内外发展趋势等[5]，通常以学科、行业、领域等为分类方法，本文统称为“学科共识”。

②专家经验

设定评价指标体系时，一般要综合考虑主导方的评价目的和评价涉及的学科共识[6]；具体评价过程则是专家对被评方之间某些方面做比对、推理和重要性排序，对所用指标做权重衡量等等；评价的结果以每位专家打分或排序为呈现形式。整个指标比对、赋权打分过程与主导方、评价方和被评方的个人的学识、理解、体会、人格特质、特殊偏好、独特感受、冷门经验等密不可分，这些知识被解构、澄清、理顺、重组甚至再创造，本文将这些个性化的知识统称为“专家经验”。

③派系立场

学科共识和专家经验虽是影响评判结果的重要依据，但在圆桌讨论、现场争辩、网上点评、打分投票时[7]，在思想交锋、知识比拼或逻辑推理等争论背后起

5 也会涉及个人（年龄、属地）和社会属性（师承关系、舆论影响）等，但属次要矛盾。

6 除了评价主导方意见（明确要求、底线、约束条件、参考标准、评价指南、评选倾向等），设定评价指标通常会考虑评价方内部的专家意见，有时也考虑被评方意见，召开专门制定评价标准的专家研讨会，请多方代表共同参与表达意见；此外，有时请被评方陈述后，专家再评审讨论，标准和指标可能会边讨论边明确。本文通盘考虑，做抽象和简化。

7 有很多评价工作并不包含讨论或辩论环节，仅打分或直接投票，以确保专家意见的独立性（如鉴定会、结题会）；也有评价工作虽然讨论但不投票，仅给出专家意见，而结果由评价主导方自行判断（如论证会、竞标会）；还有的要多轮讨论（如论文投稿评审）；有的先独立打分作初筛，再集体讨论得排名（如会议评审、评奖会）；也有既讨论也打分投票的评价

压倒性、决定性作用的因素，往往不是学科共识（争论之处已无共识、都有道理）或专家经验（个人偏好暴露时已不可信），而是学科之下、个人之上的分派、分群的人际斗争，是成体系、成建制的群体间、集团化的利益博弈和共同体范式[31][32]。这些符合派系利益博弈的、系统化的主张、观点、诉求、策略、范式、“潜规则”“圈子”、亚文化等知识，本文统称为“派系立场”。

表 1. 同行评议的知识

特征	举例	来源	稳定性	权属
学科共识	公理定理、学术常识、基本理论、实验规范、工程惯例、行业历史和前沿热点、国内外发展趋势等	组织化的（学术共同体）、非个人的；外显的、客观可查验的	在实操中表现为按照学科类别、学科层级来组建专家库、挑适合的人选；在模型中表现为专家匹配度是专家遴选多目标函数求最优解的最主要变量，调整该变量能在整体上协调每位专家工作量、专家人数等变量，而匹配度则以相对稳定的学科分级为标准。	公域
专家经验	个人的学识、经验理解、体会，甚至是特殊偏好、独特感受、冷门案例等	往往隐藏在专家头脑中、来自长期记忆和潜意识，具有独立性，不会因外界影响而轻易改变	在实操中表现为绝大多数专家能对被评方真假、质量优劣、选题重要性、可靠性、可行性、严谨性、充分性、新颖性、创新性、经济性等做出非自相矛盾的独立判断；在模型中表现为指标权重可查验，并通过微调来达到权重在多维变化时保持一致。	私域
派系立场	符合派系利益的主张、观点、诉求，人际关系、博弈策略等	派系化、集团化、建制化的；由领导偏好、资源分配、时代需求、行业惯例、社会舆论、文化历史等多因素综合决定	在实操中表现为学科间资源供给份额、学科内派系利益分配比例、派系内尊卑排名等都变化较慢，现实中派系格局也不太容易打破；在模型中表现为初始群体意见、异议包容度等典型的派系立场变量是重要的初始设定值，对最终群体意见和投票结果产生可预见的影响。	群域

（2） 同行评议的场域

通过进一步分析可知，上述三类知识分属不同主体，对应场域、领域、权利也各自不同，主导权的竞争结果也因此不同。

①学科共识 vs 专家经验

当专家遴选与指标比对交融，学科共识和专家经验竞争主导权。在该过程中，学科共识具有框定指标范围的全局性优势，能影响专家队伍的人员构成，进而极大地影响群体意见；当然，如果专家能参与评价指标设定，则专家经验就能极大影响投票结果；但是当专家不能参与指标体系设定时，个人的合理意见若要上升为群体意见，则还需一票否决机制、少数服从多数、一致通过等机制。

②专家经验 vs 派系立场

当指标比对与问答互动交融，专家经验与派系立场竞争主导权。在该过程中，指标权重虽由专家经验来判定，但派系立场对权重、专家个人的投票选择产生重

工作（如论文答辩）；等等。本文通盘考虑各个场景，做抽象化、简化分析。

要影响。在允许事前交流研讨发言时，派系立场构建出利益博弈的小环境，进而影响专家个人的投票决定（可能不同于专家内心的打分）；在问答互动后置的评价工作中，派系立场也能通过潜在的异议包容程度来影响指标权重和评价结果，而每次评价结果都在为下一次评价设定既有的派系立场。

③学科共识 vs 派系立场

当学科共识与问答互动交融，学科共识与派系立场竞争主导权。在该过程中，组织化的学科共识规定了评价初期的指标体系，群体化的派系立场规定了异议包容程度。学科共识影响评价指标、专家构成，对评价结果具有全局性约束，而派系立场影响异议包容程度、博弈策略，对评价结果具有局部性约束，二者优势孰高孰低，往往较难判断。但无论哪种情况，单独的专家经验往往不具备主导权，除非能抵消学科共识和派系立场的主导权，比如以下几种情况：多学科交叉、跨领域问题较多，专家构成的多样性较高；派系过多、或者派系势力相当，评价环境的包容度较高；该专家是得到认可的业内权威、具有专业之外的话语权等。

（3） 同行评议的逻辑

同行评议是复杂的思维过程，但总的来看包括行为：①被评对象知识向评价方展示传递，评价方理解，涉及沟通的有效性和效率，本文统称“有效沟通”。②对基本事实、定理和理论、前提和底线等规则和基准做出鉴定和认可，本文统称“基本事实和基础知识鉴定”。③专家个人特色极强的直觉性判断，包括判断问题重要性、论据可信度、解法可行性、思想新颖性、逻辑框架和技术路线及实施可行性、计算和证明严谨程度，以及专家通过预感、灵感、顿悟、武断、猜测、“赌一把”等非理性思维做出的判断，本文参照司法和行政用语[33][34]，统称“自由裁量”，或曰“认可”。④专家们统一评价结果的过程和机制，本文统称“认同”。沟通、鉴定是评价的前提和基础，认可和认同是评价的关键。

综上，同行评议可被抽象为知识运动变化的过程，是通过沟通、鉴定、认可和认同等行为，促使具有不同权属性质的知识相互竞争主导权的过程。

3.2. 同行评议的成本理论模型

（1） 问题的凝练

在上述知识分析的基础上进一步思考。对一条知识而言，其权重（相对重要性、得到认可和使用的潜力）是与权属息息相关的：一个人的观点很难比教科书、行业手册更易被接受；一位专家言之凿凿的多年经验、最新思考和前沿技巧，比纸上谈兵更令人信服；一个学派、一个机构的观点，则具备一定社会认可度。这正是实践中习以为常却往往忽略、理论界早被提及却未被解决的问题：

——实践中，同行评议存在“测不准”问题。有人问，评价中明明有清晰的指南、确定的指标和权重，怎么会测不准？实际上，现场改分、调整权重现象很常见。因为公开的指标和权重只是表象和参考，同行评议不是专家拿固定的“尺和秤”去测量被评对象的长短轻重；而是在有效沟通前提下，专家剥离对象中的专家判断已达标的基本事实和基础知识，对剩余那些具有不确定性的知识作自由裁量，判定其可行、严谨、创新等性状。权重存乎于“心”。可见，自由裁量必导致个人误差，而更要紧的是同行评议存在系统误差——很多同行评议不仅没有设置针对自由裁量误差的对冲机制，更缺乏保障评价各方有效沟通的疏通机制，漠视评价方与被评方、专家内部的充分交流，让真实观点憋在人们心中，压抑后扭曲。诚然，有多种外部干预手段能让“测不准”的评价结果得到认可，但却少

有举措正视和直面权重在“悄悄改变”、不因人为规定而保持不变的客观事实。

——理论上，什么是知识？同行评议能回答吗？知识源于大自然，也被社会构建[35]，在私人占有时赋予个体以理性，群体占有时赋予集团以利益，逐渐变为人类认识和改造自然、主客相符的共有理性，最终回归大众常识。在周而复始的过程中，正是物勒工名、专利评审、同行评议等制度[10]使得可变的知识权属明晰化，使得可变的知识使用价值明晰化。同行评议不仅是“科学家社会组织的副产品”[36]，而是“科学活动本性”之必然产物[37]，是社会对知识进行“拉扯、揉捏”、重塑和淘汰的机制。然而，什么是知识？目前还争论不休，处理知识权重变化的理论也较少。虽然阿罗悖论、孔多塞悖论指出了用序数词排序代替实质内容评价的逻辑谬误[28]，层次分析法、观点动力仿真揭示了权重、观点前后不一致现象，但这些研究较少触及权重变化的内在机理。

可见，知识属性主导权（权属）变化与其使用价值（权重）变化有逻辑上的必然联系，这是认识同行评议的重要角度——同行评议是知识的确权定价机制。

知识如何被确权定价？千百年来人类经过不断探索才形成了物价物权、股价股权等市场机制及其产权制度、立法司法体系；而跟实物商品相比，知识产权、知识价值、知识交易等体系的建立其实是近几百年才有的，很多机制并不完善。虽然有著作权、版权、专利权等知识获得权利的方式，但需要中介物、代理机制、侵权发生才能确权维权[38]，且权利边界随着事态变化也发生变化，在实现权利时还带有被动性、滞后性、外部溢出性。因此，知识的权属、权重及二者的变化，虽然理论上存在联动，但实践中确权和定价代价较高[39]。反思同行评议中的变量：

（1）可观测的量。评审时长、专家人数、评审指标、排名结果、专家背景资料、被评对象材料、会议记录、专家评语、被评人答辩等文字材料等；

（2）很难观测但终究有工具或办法能观测的量。评审中人脑的物质能量变化、评审后被评对象的发展情况、实施效果、产生的经济和社会效益等；

（3）既很难观测，又没有合适工具去观测的量。知识权属权重变化、被评对象的新颖、可靠、严谨、深刻等性状。

可见，同行评议本身带有模糊性和多样性，它只是对知识做阶段性确权定价。一方面，在实践中，同行评议的知识确权功能从属于纪法和行政工作，带有道德和习俗色彩；而其定价功能虽然表现为“排名打分”“调整权重比例”等权重设定和变化现象，但这些发生在不同的评价活动中，并不适合做横向比较，也没有综合的、跨领域或学科的知识定价统一市场。另一方面，在理论上，知识使用价值是逐步展开的，是要同使用它的人紧密联系的，同行评议阶段的知识确权只是赋予生产或发现该知识的人以最基本的被认可权，而该知识得到认可后如何被应用（如获得专利权、版权、股权等）、在应用中又创造出多少价值，则规定了该知识的更多价值。故知识的确权和定价虽然同时发生，但随着知识使用价值的变化，权利范围也随即变化。显然，确权是知识定价的必要条件，但非充分条件。

综上，本文认为在同行评议中，被可靠记录的专家评语（含评价方的质疑与被评方的答复等内容）是能收集到的且与评价内容直接相关的有效变量（必要条件），表征知识确权；而专家心中对知识权重做调整变化进而认可或否认权属，则是并不容易观测的动因变量（充分条件），表征知识定价。如果要问一个在现阶段可实证的问题，且排除权力、金钱、私情等因素干扰，单从学术、知识角度来考察确权和定价之间的必要非充分关系，则该问题是：

定价（内心认可）与确权（外显评语）的联动关系是怎样的？

（2） 假设及演绎

本文以评价中普遍存在的显然性为例，考察专家的认可行为，做假设和演绎。

表 2 关于同行评议的假设

假设	内容	解释	演绎
知识的先验性假设	知识并非被创造，它是先验的、被发现的，“一直就在那，在大自然当中”。	如果把知识不看做被“生产”“创造”出的，而是被反映、发现、找到的，则人类学习、理解和创造知识的过程可被视为连接建成的过程，包括发现可能存在的连接、尝试路径、选择出可靠可行的高效的“道路”等。实际上，评委在评审中都不得不做“事后诸葛亮”，先验性有现实性。	1. 放下具体的知识、具体问题，把关注的焦点转到“连结”的提出、选择和建立上，则理解、认可一条“新”知，就是建立一条连接。认可与知识无关，与连接的选择和尝试有关。 2. 连接是逻辑思维本身，是一类特殊的知识；是衡量知识价值的一般等价物。
知识的连续性假设	任意两条知识之间，要么是“无缝衔接”（二者可合成整体），要么存在至少一条知识或连接。	知识有简单、复杂之分，复杂知识由简单知识逐步积累叠加、交织组合、延申拓展而形成。虽然存在跳跃性、突变性判断（先验的、证据不足的武断、猜测、联想等），但不存在无中生有的知识，即便是表述上很简单的公理也是由思考复杂知识才能得到和认可的。	3. 掌握复杂知识要以掌握那些构成和支撑该复杂知识的相对简单的知识为必要非充分条件。 4. 知识可以被视为点集，可以建立数轴上的一一对应关系；而连接则可以被视为向量。 5. 知识间连接存在性是第一性，不需证明，但连接的合理性需要证明；否定连接合理需要证明。
知识的可测性假设	只要措施得当，所有知识都可被挖掘、记录、计量等。	适当措施包括但不限于：独立思考、问答辩论、数理统计、录音摄像、打分排序、公开承诺、匿名评语、开放平台、脑成像技术、名利报酬等。	6. 测量措施、成本不同，反映认可的形式也就不同。 7. 知识间连接，可知则可测；可测则可达。 8. 世上知识本没有难易之分，教学和研究的人多了，也就简单了。

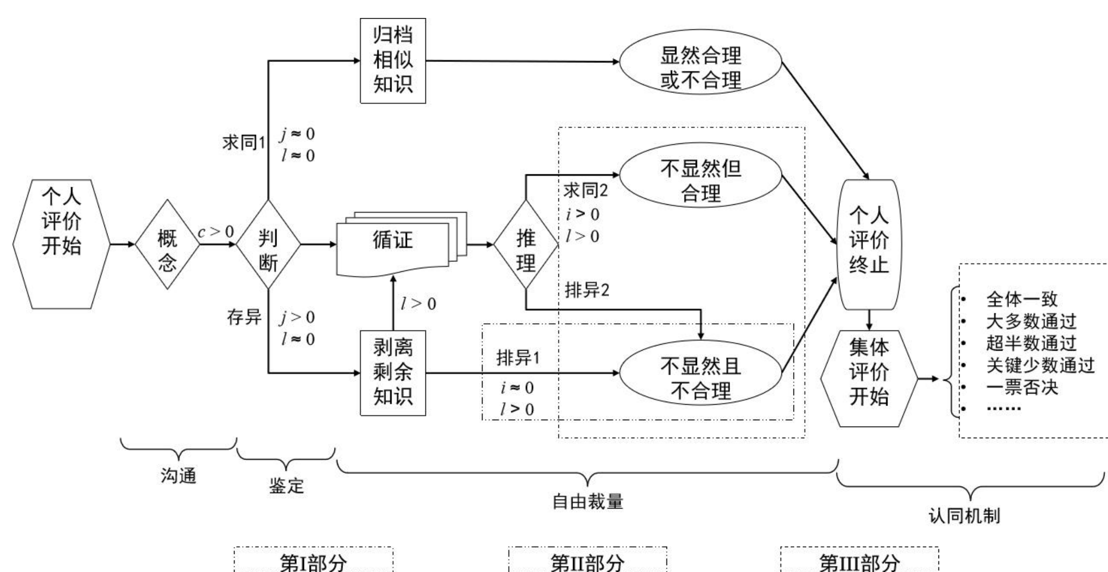


图 2. 以评价显然性为例的同行评议认可过程

以第 I 部分为例，讨论当专家认为被评对象不显然且不合理时（图 2）：

- 对任意两条知识，判断连接合理与否的成本是个变量。设判断二知识间连接不合理的难度为排异推理成本（图 2 排异 1），记为 l
- 设单位时间内的排异推理成本为排异推理强度，即排异推理力，记为 d，有：

$$d \propto \frac{l}{t} \quad (式\ 1)$$

试用（式 1）描述学习和考试过程：若待学（考）知识难度相同（l 不变），则某人排异推理力越大，其学（考）完该知识所需时间就越少；若某段时间内某人的排异推理力保持稳定（d 不变），则知识越难，学（考）完所需时间越久。

- 考察排异推理力的动态变化过程，若成本不变，而此刻排异推理力与下一刻不同，这是排异推理力的“加速度”，即敏与讷，记为 N，有：

$$N \propto \frac{d}{t} \propto \frac{l}{t^2} \quad (式\ 2)$$

试用（式 2）描述评价中的不显然性或不显见性（Non-obviousness）：是否显然（Obvious or not）是鉴别论文、专利等对象是否具备创造性、新颖性的关键概念，是对某知识具有较高使用价值的认可。若某人推理力加速（敏），则完成推理时间迅速缩短，思维跳跃，此即显然性的思维过程；反之，若推理力减速（讷），则建立连接的时间拉长，思维的由此及彼就变得滞涩，此即并非显而易见。当排异推理难度不变时（l 不变），推理时间短则显然性大，推理时间长则显然性小。

小结 I：不显然性与推理的合理性有关，而与个体差异、知识本身无关；换句话说，是否能迅速做出由 A 到 B 合理与否的推理，与张三李四无关（不需要是小同行），与 A 或 B 也无关，只与 A 到 B 之间的推理和证明的结构、顺序、力度等有关，与这个连接场域有关。在排除不合理、做出否定性推理的过程中，不显然性随时间增加而减少，趋势上大致是先骤后缓。

以第 II 部分为例，当专家开始了解被评对象、对连接存在可能性做研判时[8]：

- 不同个体对同一学科共识在表达方式上存在微小差异，在掌握学科共识的多寡和重要性排序上也存在差异，需要通过沟通对共识知识做概念界定、拆解、剥离、厘清和化归。沟通行为的成本是不得不考虑的变量，设沟通成本为 c
- 对任意知识，设识别学科共识、排除不符合要求的连接的行为需要不同成本，记鉴定出相似知识（存同）和鉴别出剩余知识（存异）的成本分别为：

求同 1 时，$j > 0, l \approx 0$ （式 3.1）

存异时，$j > 0, l \approx 0$ （式 3.2）

求同 2 和排异 2 时，$i > 0, l > 0$ （式 3.3）

试用（式 3.1）描述“求同 1”或“熟视无睹”的过程：当被评对象与学科共识关系被识别为相似或不相似时（即知识的结构或特征判别为相似或相同时），有判断成本，而推理成本很小，此过程可称为“知识归档”。

试用（式 3.2–3.3）描述“存异”或“脱颖而出”的过程：弃相似、留不同，完成“知识剥离”，对剩余知识做鉴别。就新颖性而言，当评价方接收到了被评对象对剩余知识构建连接路数的阐释或暗示时，便先要判断其是否符合科学规律、评价要求，若发现不合理，可直接否定，则时第 I 部分（排异 1）讨论的内容，

8 由于评价主体是专家而不是被评方，故我们不考察被评对象的知识传达效果（清晰与否、是否“对专家路子”等），而是考察专家的思考过程。

不显然且不合理；反之，若发现有一定合理性，但还是不显然，则要拨开杂草、寻找证据、设置目标，并依据证明效力做出判断和推理。这是循证过程，过程中若发现不合理（排异 2），则不认可；反之，是求同 2，则认可不显然但合理。

小结 II：知识间连接的合理性与显然性并不对称，合理的未必显然，不显然的未必不合理。对“不显然且不合理”的推理只是否定性的顺延（排异 2）；对“不显然但合理”的推理（求同 2）是在不同中发现相同。如果说排异 1 和 2 的主要区别是循证——否定顺延需要额外的成本，那么排异 2 和求同 1 的主要区别则是信心和包容——达到双重否定的过程中，不仅有循证排除的不显然且不合理，而且要有在信心和包容推动下新增设的知识和连接，此即授信成本。因此，循证及授信是不显然但合理的必要成本[9]。可见，所谓新颖性，实乃一个待证状态、一个构建过程。说一条知识新颖，意味着“此时”其合理性需要额外的证据来支撑，待佐证积累到可信程度时，合理性对证据的需求便达到边界，再增多证据也不会对信心增加产生较大影响，合理性也就失去不显然性，“彼时”新颖性随即消失。故新颖性是个历史范畴、一个时间函数。知识存在合理性与显然性的不对称关系可以通过“时间”这个变量来消弭：只要时间足够久，不论是不是专家，任意知识的不显然性都趋于零，不存在不显然的知识，正所谓“太阳底下无新事”；在大写的“人”面前，对所有知识而言，“我们必须知道，我们必将知道”。

限于文章篇幅，本文对显然性之外的评价及图 2 的其他部分暂不做讨论。

（3） 推论的验证

数据：由上文假设和演绎，本文提出，连接是认可的重要特征，个体差异或知识内容反而不那么相关。故对专家评语的考察以关联性词语为主，包括连词、副词、介词等，这些词语构建起知识间的连接。

数据收集依据：本文把内心认可行为置于期刊论文评审实践中，分两种情况进行讨论：一种是初审即被退稿的不认可（真阳性），另一种是稿件最终被录用但初审时尚未完全认可（假阳性）。前者是不显然且不合理，否定性评语应该随时间增加而呈幂律分布；后者是不显然但合理，但合理性尚未通过循证和授信得以实现，现阶段也应该服从幂律分布。不像多轮评审可以明确考察时间变化与评语的关系，仅有一轮的初审评语不易直接体现时间变化，故用评价行为发展过程来代替。本文设定，不认可行为大致遵循先质疑、再循证、最后推翻的发展过程。

数据来源：研究同行评议评语的文献仍然较少，研究角度无外乎内容分析[40]-[42]、文本分析[43]、词频分布规律等[44][45]。本文数据一方面是由《心理学报》编辑部邱炳武老师鼎力支持的论文初审退稿专家评语，另一方面是随机收集的《心理学报》网上公开的已录用论文评语，数量上并不多，前者 30 篇，后者 3 篇。对所有评语做词频分析，然后取简单平均，得到单篇论文评语的词频。

数据编码：在参考相关文献[46]-[51]基础上，本文的编码是遴选出否定性副词、连词、介词，并按照质疑、循证、推翻三个阶段分类，具体方案见表 3。

分析工具：本文采用 NVivo 软件，辅助评语文本的质性数据分析[52][53]。

表 3 评语编码方案

阶段	编码词语	备注
质疑	是否，是不是；为什么，为何；进一步，继续；但，但	初步否定

9 循证成本的排异 2 可称为推翻成本，与排异推理应该是结合（相乘）的关系；求同 2 可称为授信成本，它相对复杂，与创造力、联想力等有关，本文不展开讨论。

	是，然而，却；不一定，不见得，未必，没准，说不定，非必然	
循证	假如，如果，假设，或者；尽管，否则	否定的等待期
推翻	不可能，绝非，绝不，绝未，决然，决计；并非，从未，从来，并未，并不，并没，丝毫，断非，断不，断断，断乎，断然，万万，全然，压根儿，切莫；何以，怎么就，怎能，白白，枉费，白费，徒劳，空费	否定顺延则推翻；循证未完则并非完全推翻

分析结果：关联词词频在三阶段过程中，的确呈现幂律分布，真阳性总体成本低于假阳性，假阳性的质疑和循证成本明显高于真阳性。可见，评审成本在被退稿和待复审之间，存在较大差异，很明显是循证成本，此时授信成本还不明显。

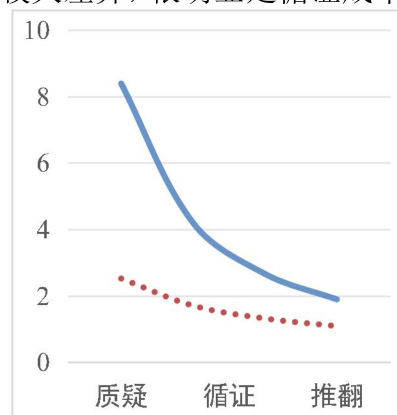


图 3. 专家评语词频分析（均值）：蓝实线为假阳性，黄虚线为真阳性

这一结果符合常识和本文模型预期，但幂律分布也可能由其他因素导致或是本文条件约束的结果（如编码方案、聚焦显然性、评语文字量等），何况数据量还是太小。因此，需要做进一步预测和验证：下面纳入作者答复再做统计，答复中肯定会含有对评语的否定，故词频应该是整体增长的，但比起首轮评审，多轮评审应当体现出显然性降低；成本变化趋势应当趋升，即曲线降势趋缓。

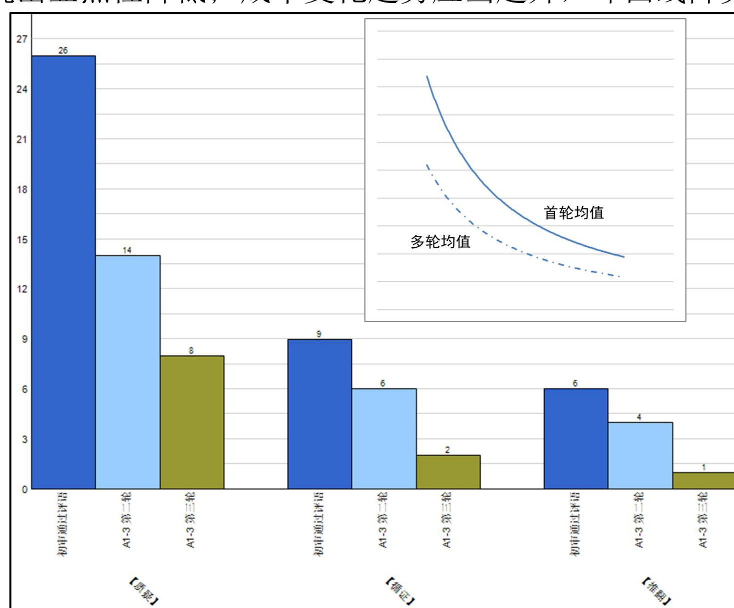


图 4. 专家评语词频分析：多轮比首轮的显然性低、降势趋缓

词频结果符合本文预测，多轮评审的显然性低于首轮；其降势比首轮趋缓，意味着循证和授信成本的投入。未来如能收集更多数据，应做更多检验和社会科学化淬炼，若找到反例，则本文的假设或演绎就需要修改或者抛弃。

综上，知识确权（内心认可）与定价（外显评语）在同一时空场域中被联立起来，理论模型暂时自圆其说，但可被推翻。此即本文对同行评议确权定价内在机理做出的解释——同行评议的成本理论模型。

3.3. 同行评议公理化体系构建

（1） 定义

1. 评价：可以用构建映射（连接）来抽象表示评价，即，评价是构建映射。
2. 评价的前提：物质能量的绝对运动是一切的源头。随机无序运动的物质能量偶然重复出现，这种重复出现（独立的自身、秩序）不同于“其他”（环境）。本文把重复出现、保持不变的物质能量称为“有序”、“质能体”或“自由意志”。有序出现即是二元对立产生，形成质能体与非质能体，亦即“质能体-环境”。二元对立的双方彼此反映，成为对方的映射基础，评价有了发生的前提。
3. 评价的动力：二元对立、彼此反映为评价奠定基础，而绝对运动必然让“反映”与“自身”存在差异，该差异为构建映射（评价）提供了最初动力源。差异引发映射的构建，推动对立向统一转化，这便是评价的动力。
4. 评价的测量：可以用时间来测量质能体的连绵发展（连续性），用空间来测量质能体的内容和程度（复杂度）。
5. 评价的理论单元：“质能体-环境”是评价的抽象化基本单元。
6. 评价的最小单元：设想质能体仅发生极其微小的运动变化，通过时间来对此测量，则时间的累积就是最小的评价；设想质能体的运动变化仅发生在极其微小的时间，通过秩序来对此测量，则秩序的累积就是最小的评价。可见，评价的最小单元可以是纯时间的累积，也可以是纯秩序的累积。
7. 评价的复杂单元：称达到一定复杂度的质能体为知识，称保持一定连续性的知识为脑，称脑的自映射为意识。故评价和意识不可分割。有意识的质能体就是智能体。智能体是评价的复杂单元。
8. 评价的现实单元：“脑-环境”是评价的实体化基本单元。
9. 评价单元的一般情况：由二元发展出多元，多元的对立统一是一般情况。

（2） 公理

公理 1：知识的先验性。知识并非被创造，它是先验的、被发现的，“一直就在那，在大自然当中”。

公理 2：知识的连续性。任意两条知识之间，要么是“无缝衔接”（二者可合成整体），要么存在至少一条知识或连接。

公理 3：知识的可测性。若措施得当，则所有知识都可被挖掘、记录、计量。

（3） 推论

除了本文表 2 的演绎推论和小结 I、II，还有如下推论：

- 评价中的权重变化，既跟知识间连接的合理性有关，也跟是否新设置知识或连接有关，这些受到专家经验和派系立场影响，故知识权属影响了知识价值；反过来，个人经验和群体偏好会在知识价值的（不）认可和认同中得到重塑。
- 权重变化背后有权属影响，这跟利益有关，也跟脑的生理基础有关。就如同视错觉一样，长短感在不同背景下会有错觉，知识价值也可能类似。

- 评价正当性来自知识，知识充足乃至溢出方有授信。评价不当首先是无知，其次是没有信心。从这个意义上看，知识就是信心。授信都是“事后诸葛亮”，与创造、学习、创新在数学上可能存在不完全对称、耦合结构和突变模型[54][55]。
- 评价是否定性权力，是自上而下的授权和自由裁量的特权，评价者就是被赋予对某条知识价值具有否决权的人，评价者和评价活动需要自下而上的监督，即对自由裁量的审查、对投票博弈的设计、对期刊分级和人才“帽子”监管。外部监督要内化为专家内省和权力让渡机制[56]。
- 评价与创新两种行为存在价值“剪刀差”，评价成本往往大幅低于创造成本，而这是评价的激励机制改革关键，是激励政策指标设计的经济学基础。
- 降低评价中循证和授信成本的方法是建立知识交换市场：让更多专家参与并推行评语公开，知识外溢最大化；同时，造成评价多方势均力敌的态势，立威慑机制。这是用空间换时间以降低成本的合理选择，是开放科学的一个理论支撑点。具体做法可参考《心理学报》和 ChinaXiv 案例[7] [55]。
- 没有评语的、不给证据的、不给答辩机会的同行评议，都是耍流氓。
- 应该根据评价的成本理论构建起一套同行评议的经济学模型，用于知识交换市场建设、开放科学的公开评审监督、创新激励机制设计、科技创新宏观指标设计、信息社会化和融资社会化研究等，进而将专家评审费和薪酬、评价的道德和伦理规范、必要劳动时间核算等建立在经济分析和实证基础上。
- 每次评价时评价者拨开迷雾、发现真知（剩余知识），都是一个善端发祥的可能。要像呵护幼崽一样呵护新知及其创造者。创新者可以只求真，评价者不能不求善。否定或认可一条新知在于“由真鉴善”，这是评价的真谛。

总之，本文在前人对评价做公理化研究的基础上[57]-[58]，借鉴传承发展，为同行评议新范式奠定基础。本文模型化和公理化的探讨还非常肤浅，谨抛砖引玉。

4. 讨论

评价是很可爱的，宜笑面应对。

当下，从论文录用、职称评审、到项目审核、机构认定，从人才筛选、资源配置，到荣誉奖励、盖棺定论等，多种多样的同行评议是对知识价值的交叉检验，但因知识创新太快、评审时间不足、被评方权利得不到保障、“四唯”特权不破、科学的投票认同机制等问题而变得满目疮痍。若不能做好评价的成本控制、没有一个成熟的知识交换、公开评审的市场，则评价的有效性可靠性会一直无法保障。这些问题裹挟在知识爆炸之中，不断发酵，由局部的、孤立的、暂时的问题，蜕变为全局的、连锁的、长期的僵局。小修小补如隔靴搔痒，局部改良似缘木求鱼。当深入调研，上下联动，以系统之观念谋划系统之变革，以创新之工具推动工具之创新，为评价立新标，还评价之初心：沟通、改进、信任、包容、呵护。

“月光下，有一颗小小的蛋，躺在叶子上……啵的一声，蛋开了，飞出一只花蝴蝶”。同行评议作为评价制度的一种，我国古代早已有之，只是形式不同、侧重点不同罢了，如井田制下农田鉴定、贤均从众、物勒工名、相地衰征、科举制度、针入水出等等，这些制度从古至今生生不息，犹如蚕蛹化蝶，虽今非昔比，但却是我们评价行为的血脉、灵魂，是中华文明自古传承的价值思想的根。今天，同行评议制度是历史范畴和国家战略，需与时俱进、独立自主。特别是在交叉创新突飞猛进、人工智能（Artificial Intelligence, AI）技术应用愈加广泛的当下[59]-[71]，我们将面临全新的人工审核、评价、监管、治理挑战，同行评议既是受益者也是

被改造者，其中蕴含大量机会。换个角度看，同行评议实际上跟具体的人、知识可能都没关系，它是一个时代、一个社会的知识观、人才观、进取心和斗争意识。

创新多非议，改革鲜平川。真正阻碍评价改革的，还是“什么是知识”这一无解的老问题。然而，我们拥有千年文明经验，有信心和决心在同行评议体制机制改革创新中实现新旧评价制度的传承发展，把传统与现代共通的精髓进行铆接、嵌合，用多种方法引古流推今潮，注新泉润老潭，让古老的中华文化涌动出全新的波浪。当然，评价不仅是中国的，也是世界的，是全人类的，评价本身就是各色知识的“过滤器”“大熔炉”，具有开放性和包容性。同行评议改革创新还要顾及多元价值和不同文明差异，在实践中不断打磨完善，在创造中逐渐形成具有中华文化底色和自主知识体系的中国式现代化同行评议新范式。

5. 结语

本文运用介尺度分析工具对同行评议做贯通式整合，同行评议是知识运动变化的过程，不同权属性质的知识在沟通、鉴定、认可和认同过程中相互竞争主导权，最终得到确权定价；由此，本文构建了显然性评价的成本理论模型，发现评价者的内心认可与外显评语间可能存在幂律分布规律，通过对 30 篇初审退稿评语和 3 篇录用论文三轮评审意见做质性数据分析，初步验证了推论，最后提出一套包含 9 条定义和 3 条公理的同行评议的公理化体系。

天边泛起鱼肚白。评价制度“破立并举、先立后破”是一项系统工程、一项伟大事业。教育科技人才一体推进探索中国式现代化的同行评议新范式，是推进高水平科技自立自强的必由之路，也必将成为提升国家创新体系整体效能的点睛之笔。“梦虽遥，追则能达；愿虽艰，持则可圆”。

致谢

本研究之所以能获得宝贵的论文专家评语，得益于《心理学报》、ChinaXiv 等期刊平台长期以来对开放科学、公开评审工作的坚守和笃行。衷心感谢邱炳武的宝贵一手资料和耐心指导；感谢张学新在公开评审、对分课堂等方面做的指导；本文还学习参考了前人研究成果，特别是张光耀、王贤文等老师的研究数据及其慷慨提供的未曾发表的数据分析结果（其收集的 888 篇材料是支撑本文假设的），这些研究为本文提供了重要借鉴和参考。感谢黄增阳和柳堤的指导和鼓励，感谢李响，王彦雨，李海英，滕飞，刘静羽，王玉菊，李正伟，亢宽盈，王哲然，黄磊，姜海婷，姜姗，李明洋等人的帮助。

参考文献

[1]. 党的二十届三中全会《决定》学习辅导百问编写组.党的二十届三中全会《决定》学习辅导百问[M].学习出版社,党建读物出版社,2024.8.

[2]. 党的二十大报告学习辅导百问编写组.党的二十大报告学习辅导百问[M].党建读物出版社,学习出版社,2022.10.

[3]. 《中共中央关于进一步全面深化改革、推进中国式现代化的决定》辅导读本编写组.《中共中央关于进一步全面深化改革、推进中国式现代化的决定》辅导读本[M].人民出版社,2024.7.

[4]. 人民网 习近平系列重要讲话数据库. 全国科技大会国家科学技术奖励大会两院院士大会在京召开. [OL]. (2024-06-25)[2025-01-09]. https://jhsjk.people.cn/article/40263596.

[5]. 新华社. 习近平在中共中央政治局第三次集体学习时强调 切实加强基础研究 夯实科

技自立自强根基. [OL]. (2023-02-22)[2025-01-09].
https://www.news.cn/politics/leaders/2023-02/22/c_1129386597.htm.
[6]. 人民网 习近平系列重要讲话数据库. 习近平在省部级主要领导干部学习贯彻党的二十届三中全会精神专题研讨班开班式上发表重要讲话. [OL]. (2024-10-29)[2025-01-09]. https://jhsjk.people.cn/article/40349781.
[7]. 李政. 于无声处听惊雷——两则十年探索案例对“破四唯、立新标”的启示.[R]三思派,2023 年 11 月 14 日.
[8]. 张耀铭.挑战与进路:数字时代的同行评议[J].理论与改革,2023(3):24-42.
[9]. 廖苗,闫曦月.后常规之后:开放科学会成为一种新范式吗?[J].科学学与科学技术管理, 2023, 44(4):21-37.
[10]. 李政,肖小溪,韩军徽,黄雷,翟岩.破“四唯”背景下浅谈同行评议新范式[J].今日科苑,2022,6:29-41.
[11]. 李静海,胡英,袁权.探索介尺度科学：从新角度审视老问题[J]. 中国科学：化学，2014,44(03):277-281.
[12]. Batterham R. Compromise through competition: a more widely application approach? [J]. Engineering, 2016(02): 286-287.
[13]. Li J., Huang W., Chen J. Possible roadmap to advancing the knowledge system and tackling challenges from complexity [J]. Chemical Engineering Science, 237(2021)116548:1-6.
[14]. 柳堤,李政.遵从科创规律：再谈科创成果与年龄的关系[J].今日科苑,2021,6:17-26.
[15]. 张素琪.专家分配问题的启发式求解方法研究[D].河北工业大学硕士学位论文,2006.
[16]. 黄亚明.神经网络在基础研究成果综合评价中的应用[D].沈阳:中国医科大学，2004.
[17]. Thomas L. Saaty. The Analytic Hierarchy Process[M].New York:McGraw-Hill Inc,1980.
[18]. 许树柏.层次分析方法中一种新的动态排序模型[J].系统工程学报,1986:42-55.
[19]. 张晋安,张汝梁,段宜武,吴为平.教育测评 A-FA 模型研究[J].教育评估,1989(3):67-73.
[20]. 俞校明,张维全.科研成果评价的模糊综合评价法[J].甘肃科学学报,2000,12(2):85-88.
[21]. 冯学军,邹杰涛.A-AF 模型在科研成果评价中的应用[J].长春光学精密机械学院学报,1999,22(1):58-61.
[22]. GB-T44143-2024.科技人才评价规范. [S]. 全国标准信息公共服务平台. 2024.
[23]. 焦艺茹. 主动观点动力学模型——建模主动交互下的群体观点演化[D]. 哈尔滨：哈尔滨工业大学，2020.
[24]. 杜姝玥.融入情感因素的网络舆情热度预测研究——以暴雨洪涝类突发事件为例[D].中南大学硕士学位论文,2023,6.
[25]. 申子文.带有多个相互依赖关系的观念动力学模型[D].安徽工程大学硕士学位论文,2023,6.
[26]. 孙润利.群体压力作用下微博网络舆情传播演化模型[D].桂林电子科技大学硕士学位论文,2023,5.
[27]. 艾凯丽.基于情感动员的社交网络观点演化研究[D].华中科技大学硕士学位论文,2021,5.
[28]. [美]兰维尔 Langville, A.N,[美] 梅耶 Meyer, C.D.谁排第一?关于评价和排序的科学[M].机械工业出版社,2014.
[29]. 中国科学技术协会,国际科学、技术与医学出版商协会.中国开放获取出版发展报告2022[R].北京:科学出版社,2023.
[30]. 亢列梅,杜秀杰,荆树蓉,等.开放科学和科研评价改革背景下我国学术期刊同行评议的改革趋向[J].编辑学报, 2021, 33(6):615-619.
[31]. 吴述尧.同行评议方法论[M].北京:科学出版社,1996:11.
[32]. 郭碧坚,韩宇.同行评议制——方法、理论、功能、指标[J].科学学研究, 1994,12(8):63-73.
[33]. 赵芳.法官自由裁量权及其理性控制[D].山东大学硕士学位论文,2006.
[34]. 胡泓.行政自由裁量权研究[D].苏州大学博士学位论文,2006.
[35]. 李正风.知识,创新与国家创新体系[J].山东科技大学学报:社会科学版, 2011, 13(1):18-24.

[36]. 默顿.鲁旭东,林聚任译.科学社会学（上、下）[M].北京:商务印书馆,2016.
[37]. 巴恩斯.尼科尔斯编.魏洪钟译.托马斯 库恩和科学中的社会秩序问题.[C].上海复旦大学出版社,2013.
[38]. 张五常.经济解释[M].中信出版社,2015.
[39]. 师萍,郭杰.科技税收优惠与纳税筹划[M].科学出版社,2010.
[40]. Paltridge B.Front Pages.The Discourse of Peer Review, Reviewing Submissions to Academic Journals[M].London: Palgrave Macmillan London, 2017.DOI:10.1057/978-1-137-48736-0.
[41]. 牛田雨.《同行评议话语》（节选）翻译实践报告[D].重庆大学硕士学位论文,2021.
[42]. 张朝霞.法学博硕学位论文评议失当的问题分析与规制建议[J].新文科教育研究, 2021(3):127-139.
[43]. 张光耀,谢维熙,夏鑫璐,耿屿,姜春林,王贤文.非共识研究的影响力更高还是更低呢?——基于中文开放同行评议审稿意见的分析[J].情报理论与实践,2022,9:102-109.
[44]. Zipf, G. K. Human behaviour and the principle of Least Effort: An introduction to human ecology [M].Cambridge, MA: Addison Wesley, 1949.
[45]. Mazzarisi, O., de Azevedo-Lopes, A., Arenzon, J. J., Corberi, F. Maximal diversity and Zipf's law[M].Phys. Rev. Lett, 127: 128301,2021.
[46]. 张蕾.概念结构及其应用[D].西北工业大学博士学位论文,2001.
[47]. 胡曼媛.含有否定性成分的或然类词语研究[D].华东师范大学硕士学位论文,2018.
[48]. 艾贵金.近代汉语五类语用标记研究[D].上海大学博士学位论文,2018.
[49]. 李明敏.汉语否定性情态副词研究[D].华侨大学硕士学位论文,2011.
[50]. 石毓智.现代汉语的否定性成分[J].语言研究, 1989(2):12-21.
[51]. 孙琴.现代汉语否定性结构专用副词的考察[D].广西师范大学硕士学位论文,2005.
[52]. 李政,罗晖,李正风,等.基于质性数据分析的中美创新政策比较研究——以“中国双创”与“创业美国”为例[J].中国软科学, 2018(4):13.
[53]. 李政.基于质性数据分析的社会学研究方法:理论与案例[M].中国科学技术出版社,2022.
[54]. 柳堤、李政、孙智源、李同飞.定性与定量视角下的学习观综述[J].今日科苑, 2020(8):77-85.
[55]. 李政,罗晖,李正风,王宏伟.基于突变理论的科技评价方法初探[J]. 科研管理,2017(S1):193-200.
[56]. 李正风,阎妍,武晨箫.学术评价破"五唯"与科学共同体的责任[J].科学学与科学技术管理, 2022, 43(10):17-28.
[57]. 袁劼.基础评价理论的公理化分析与构建研究[D].上海交通大学硕士学位论文,2009.
[58]. 孟广武.公理化方法及其在社会科学中的应用[J].聊城大学学报：自然科学版, 2013, 26(3):1-6.
[59]. 任一晗.人工智能算法司法应用问题研究[D]东北林业大学硕士学位论文. 2023.
[60]. 魔王,杜伟.AI 自动评审论文，CMU 这个工具可行吗？我们用它评审了下 Transformer 论文[N].澎湃新闻,2021-02-04.
[61]. 网易号 Justscience. AI 在同行评审中的应用：优势与风险 [OL]. (2023-09-28)[2025-01-09]. www.163.com/dy/article/IFNTMQJC0532AN5N.html.
[62]. 搜狐 时光新文. AI 参与论文同行评议，又一项工作可以“托管”了？[OL]. (2024-05-21)[2025-01-09]. https://www.sohu.com/a/780546425_121955986.
[63]. 张彤, 尹欢, 苏磊, 王静, 夏道家. 人工智能辅助学术同行评议的应用及分类[J]. 中国科技期刊研究, 2021, 32(1): 65-74.
[64]. 徐亦舒,查欣雨. 人工智能辅助同行评议的技术潜力、应用领域与伦理隐忧[J]. 数字出版研究,2023,2(3):21-27.
[65]. A Critical Examination of the Ethics of AI-Mediated Peer Review. [OL]. (2023-09-02)[2025-01-09]. https://arxiv.org/abs/2309.12356.
[66]. Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews. [OL]. (2024-03-11) [2025-01-09].

https://arxiv.org/abs/2403.07183.
[67]. Kankanhalli, Atreyi. Peer review in the age of Generative AI [J]. Journal of the Association for Information Systems, 2024, 25(1), 76-84.
[68]. Checco, A., Bracciale, L., Loreti, P. et al. AI-assisted peer review. Humanit Soc Sci Commun 8, 25 (2021).
[69]. The advent of human-assisted peer review by AI. Nat. Biomed. Eng, 2024, 8:665–666.
[70]. 叶继元,郭卫兵.生成式人工智能参与学术评价的反思[J].中国社会科学评价, 2024(1):37-48.
[71]. Anagha Nair. Improving Peer Review With Technological Innovations: A comparative analysis of 6 AI tools. Enago Academy. September 24, 2024. https://www.enago.com/academy/6-ai-tools-peer-review-process/.